PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-293764
(43)Date of publication of application : 04.11.1998

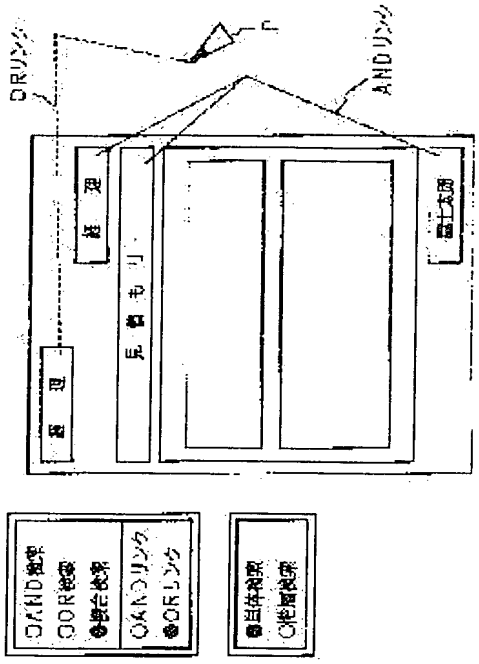
(51)Int.Cl. G06F 17/30

(21)Application number : 09-102159
(22)Date of filing : 18.04.1997

(71)Applicant : FUJITSU LTD
(72)Inventor : GOTO MASATOMO

(54) STRUCTURED DOCUMENT DATABASE RETRIEVING METHOD, STRUCTURED DOCUMENT DATABASE RETRIEVAL SYSTEM AND RECORDING MEDIUM

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a retrieving method of a structured document database which easily designates an element of an object to be retrieved, easily inputs a keyword to the element and performs retrieval in a short time even at the time of retrieving a structured document that is unknown to a user in a retrieving method to a database which accumulates element names that constitute document and structured documents in which the hierarchical structure of an element is defined.
SOLUTION: This device shows a document format diagram that defines layout positions of contents of an element, accepts an input of a keyword that is associated with the layout position of the document format diagram, identifies an element that is an object to be retrieved through the layout position where a keyword in inputted, retrieves the content of the identified element based on an inputted keyword and acquires a document.



LEGAL STATUS

[Date of request for examination] 13.03.2001
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平10－293764

(43)公開日　平成10年(1998)11月４日

(51)Int.Cl.[6]	識別記号	ＦＩ		
Ｇ０６Ｆ　17/30		Ｇ０６Ｆ	15/403	３２０Ａ
			15/40	３７０Ａ
			15/403	３２０Ｚ

審査請求　未請求　請求項の数13　ＯＬ　（全 12 頁）

(21)出願番号　特願平9－102159

(22)出願日　平成９年(1997)４月18日

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号

(72)発明者　後藤　正智
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

(74)代理人　弁理士　河野　登夫

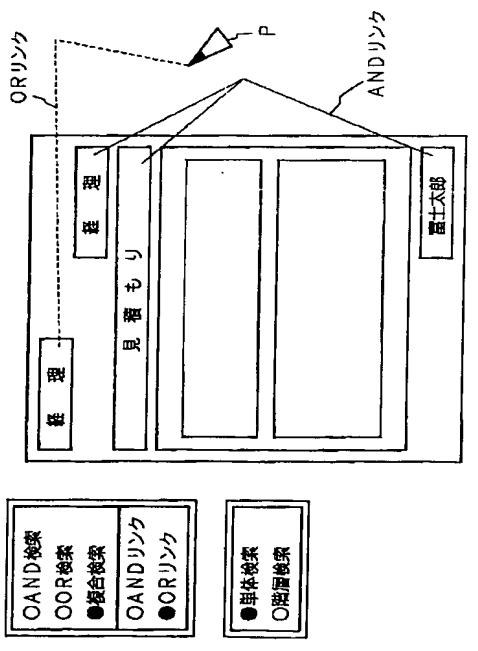
(54)【発明の名称】　構造化文書データベース検索方法、構造化文書データベース検索システム及び記録媒体

(57)【要約】

【課題】　SGML (Standard Generalized Markup Language :ISO 8879)で記述された文書のように、文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が蓄積されているデータベースに対する検索方法に関し、文書構造がユーザにとって未知である構造化文書を検索する場合であっても、検索対象のエレメントの指定、及びそのエレメントに対するキーワードの入力が容易であって、検索が短時間で行える構造化文書データベースの検索方法を提供する。

【解決手段】　エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示し、文書フォーマット図のレイアウト位置に対応付けたキーワードの入力を受け付け、キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別し、入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得する。

スタイル情報を用いたキーワード入力例の図



1

【特許請求の範囲】

【請求項１】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義され、文書におけるエレメントの内容の表示方法を規定する文書の成形規則が関連付けられた構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対する検索方法において、

エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示し、

文書フォーマット図の前記レイアウト位置に対応付けたキーワードの入力を受け付け、

キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別し、

入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得することを特徴とする構造化文書データベース検索方法。

【請求項２】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対する検索方法において、

所定の階層に位置するエレメント名を、その下位の階層情報の表示情報とともに表示し、

表示したエレメント名より下位の階層に位置する、エレメント名を表示すべき階層の指定を受け付け、

指定された階層のエレメント名を、その階層関係を表す情報とともに表示し、

表示したエレメント名に対応付けたキーワードの入力を受け付け、

キーワードが入力されたエレメント名によって検索対象のエレメントを識別し、

入力されたキーワードに基づき、識別したエレメントの内容を検索し、文書を獲得することを特徴とする構造化文書データベース検索方法。

【請求項３】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対する検索方法において、

検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示し、

該構造図のエレメント名に対応付けたキーワードの入力を受け付け、

キーワードが入力されたエレメント名によって検索対象のエレメントを識別し、

入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得することを特徴とする構造化文書データベース検索方法。

【請求項４】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義され、文書におけるエレメントの内容の表示方法を規定する文書の成形規則が関連

2

付けられた構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対する検索方法において、

キーワードの入力方法の指定を受け付け、

指定された入力方法に応じて、

エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示し、

文書フォーマット図の前記レイアウト位置に対応付けたキーワードの入力を受け付け、

キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別し、

又は、

所定の階層に属するエレメント名を、その階層関係を表す情報とともに表示し、

表示したエレメント名及びその階層関係を表す情報に基づく、エレメント名を表示すべき階層の指定を受け付け、

指定された階層のエレメント名を、その階層関係を表す情報とともに表示し、

表示したエレメント名に対応付けたキーワードの入力を受け付け、

キーワードが入力されたエレメント名によって検索対象のエレメントを識別し、

又は、

検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示し、

該構造図のエレメント名に対応付けたキーワードの入力を受け付け、

キーワードが入力されたエレメント名によって検索対象のエレメントを識別し、

指定された入力方法によって識別したエレメントの内容を、入力されたキーワードに基づいて検索し、文書を獲得することを特徴とする構造化文書データベース検索方法。

【請求項５】　識別したエレメントに加え、該エレメントの下位の階層に属するエレメントを検索対象とする請求項１乃至４のいずれかに記載の構造化文書データベース検索方法。

【請求項６】　検索対象のエレメントが複数の場合、ＡＮＤ検索又はＯＲ検索のいずれかの検索条件で検索を実行する請求項１乃至５のいずれかに記載の構造化文書データベース検索方法。

【請求項７】　検索対象のエレメントが複数の場合、ＡＮＤ検索及び／又はＯＲ検索の指定を受け付け、指定された検索条件で検索を実行する請求項１乃至５のいずれかに記載の構造化文書データベース検索方法。

【請求項８】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義され、文書におけるエレメントの内容の表示方法を規定する文書の成形規則が関連

3

付けられた構造化文書が構造的に分割されて蓄積されている構造化文書データベースの検索システムにおいて、
エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示する手段と、
文書フォーマット図の前記レイアウト位置に対応付けたキーワードの入力を受け付ける手段と、
キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別する手段と、
入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得する手段とを備えたことを特徴とする構造化文書データベース検索システム。

【請求項９】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースの検索システムにおいて、
所定の階層に位置するエレメント名を、その下位の階層情報の表示情報とともに表示する手段と、
表示したエレメント名より下位の階層に位置する、エレメント名を表示すべき階層の指定を受け付ける手段と、
指定された階層のエレメント名を、その階層関係を表す情報とともに表示する手段と、
表示したエレメント名に対応付けたキーワードの入力を受け付ける手段と、
キーワードが入力されたエレメント名によって検索対象のエレメントを識別する手段と、
入力されたキーワードに基づき、識別したエレメントの内容を検索し、文書を獲得する手段とを備えたことを特徴とする構造化文書データベース検索システム。

【請求項１０】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースの検索システムにおいて、
検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示する手段と、
該構造図のエレメント名に対応付けたキーワードの入力を受け付ける手段と、
キーワードが入力されたエレメント名によって検索対象のエレメントを識別する手段と、
入力されたキーワードに基づいて、識別したエレメントの内容を検索し、文書を獲得する手段とを備えたことを特徴とする構造化文書データベース検索システム。

【請求項１１】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義され、文書におけるエレメントの内容の表示方法を規定する文書の成形規則が関連付けられた構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対して検索を実行するコンピュータでの読み取りが可能なプログラムコード手段を格納しており、前記コンピュータでの使用が可能な記録媒体において、

4

前記プログラムコード手段が、
前記コンピュータに、エレメントの内容のレイアウト位置を規定した文書フォーマット図を表示させるコード手段と、
前記コンピュータに、文書フォーマット図の前記レイアウト位置に対応付けたキーワードの入力を受け付けさせるコード手段と、
前記コンピュータに、キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別させるコード手段と、
前記コンピュータに、入力されたキーワードに基づいて、識別したエレメントの内容を検索させ、文書を獲得させるコード手段とを含むことを特徴とする記録媒体。

【請求項１２】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対して検索を実行するコンピュータでの読み取りが可能なプログラムコード手段を格納しており、前記コンピュータでの使用が可能な記録媒体において、
前記プログラムコード手段が、
前記コンピュータに、所定の階層に位置するエレメント名を、その下位の階層情報の表示情報とともに表示させるコード手段と、
前記コンピュータに、表示したエレメント名より下位の階層に位置する、エレメント名を表示すべき階層の指定を受け付けさせるコード手段と、
前記コンピュータに、指定された階層のエレメント名を、その階層関係を表す情報とともに表示させるコード手段と、
前記コンピュータに、表示したエレメント名に対応付けたキーワードの入力を受け付けさせるコード手段と、
前記コンピュータに、キーワードが入力されたエレメント名によって検索対象のエレメントを識別させるコード手段と、
前記コンピュータに、入力されたキーワードに基づき、識別したエレメントの内容を検索させ、文書を獲得させるコード手段とを含むことを特徴とする記録媒体。

【請求項１３】　文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が構造的に分割されて蓄積されている構造化文書データベースに対して検索を実行するコンピュータでの読み取りが可能なプログラムコード手段を格納しており、前記コンピュータでの使用が可能な記録媒体において、
前記プログラムコード手段が、
前記コンピュータに、検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示させるコード手段と、
前記コンピュータに、該構造図のエレメント名に対応付けたキーワードの入力を受け付けさせるコード手段と、

5

前記コンピュータに、キーワードが入力されたエレメント名によって検索対象のエレメントを識別させるコード手段と、

前記コンピュータに、入力されたキーワードに基づいて、識別したエレメントの内容を検索させ、文書を獲得させるコード手段とを含むことを特徴とする記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、例えば SGML (Standard Generalized Markup Language :ISO 8879)で記述された文書のように、文書を構成するエレメントの名称、及びエレメントの階層構造が定義された構造化文書が蓄積されているデータベースに対する検索方法、検索システム及びそのデータベースに対する検索のコンピュータプログラムが記録されている記録媒体に関する。

【０００２】

【従来の技術】アメリカでは、国防総省、多くの学術研究機関がSGMLを導入しており、民生利用も拡大している。また、インターネットで人気が高いWWW (World-Wide Web)で流通しているコンテンツも、HTML (Hyper Text Markup Language) で記述された構造化文書である。また、インターネット／イントラネットの普及により、インターネットの場合は世界中の文書に、イントラネットの場合は会社のようなグループ内の文書にネットワークを通じてアクセスできるようになっている。

【０００３】特にイントラネットの普及で、社内文書のような定型文書が電子化されて蓄積されるようになり、ペーパーレス化が促進されている。アメリカ国防総省が、企業活動に関わるすべての情報を電子化してオンライン上で情報交換できるようにしようと展開しているCALS (Continuous Acquisition and Life-cycle Support) プロジェクトでは、情報交換にSGMLを採用しているので、構造化文書はより一般的になると考えられる。

【０００４】構造化文書は論理構造を持っているため、この論理構造を利用することで検索が高速にしかも容易に行える。即ち、SGML文書は文書型定義 (DTD)により文書の構造が定義され、その構造は図11に示すようなツリー構造で表される。従って、例えば、「レポート」というエレメントが「宛先」「タイトル」「本文」「担当」及び「発信元」のエレメントから構成され、これらの中の「本文」が「概要」及び「詳細」のエレメントから構成されている文書構造を持つ文書群から、「○○さん」を宛先として発信されたレポートを検索する場合、全文を検索の対象としなくても、「宛先」のエレメントを指定し、「○○さん」をキーワードとして検索すればよい。

【０００５】

【発明が解決しようとする課題】しかし、構造化文書において上述のような検索を行おうとする場合、ユーザは、キーワードを与えるときに検索対象とすべき文書構造を指定しなければならないので構造化文書の文書構造を把握しておく必要がある。

6

【０００６】例えば、ユーザは、図11のような８つのエレメントからなるツリー構造のノード毎に付けられているタグ名（エレメント名）を把握して、「○○さん」に発信されたレポートを検索する場合は、「宛先」というエレメント名が検索対象であることを指定し、「○○さん」をキーワードとして入力してやらなければならない。

【０００７】従って、このような検索手法は、高速に検索ができるにもかかわらず、構造化文書の文書構造を把握しておかなければならないという負担が大きいため、ユーザは、従来の全文検索の手法、リレーショナルデータベースの検索手法等を用いる方が検索が容易であると感じる。

【０００８】本発明はこのような問題点を解決するためになされたものであって、構造化文書の文書構造の図式をエレメント名を用いてユーザに示し、図式上のエレメント名に対してユーザがキーワードを入力したエレメントを対象として検索することにより、文書構造がユーザにとって未知である構造化文書を検索する場合であっても、検索対象のエレメントの指定、及びそのエレメントに対するキーワードの入力が容易であって、検索が短時間で行える構造化文書データベースの検索方法、検索システム及び構造化文書データベース検索のコンピュータプログラムが記録されている記録媒体の提供を目的とする。

【０００９】

【課題を解決するための手段】図１は本発明の構造化文書データベース検索方法を実施するシステムの構成図である。表示装置１は、エレメント名を用いた、エレメントの内容のレイアウト位置を規定した文書フォーマット図、構造化文書の文書構造の所定の階層の構造、又はエレメントの階層構造の構造図といった、構造化文書の文書構造の図式と図式上のエレメント名に対するキーワードの入力画面、検索対象のエレメントがキーワードを入力されたエレメントだけであるか、下位の階層のエレメントを含むかを指定し、またOR、AND の検索条件を指定する画面、検索結果をディスプレイに表示する。

【００１０】キーボード、マウス等からなる入力装置２は、画面に表示された文書構造の図式上のエレメント名に対してキーワードを入力する手段、またキーワードの入力方法を指定する手段、検索条件を指定する手段、各種入力画面の表示命令を入力する手段、及び検索の実行命令を入力する手段である。

【００１１】インタフェースモジュール３は、入力装置２を用いた、ユーザによるキーワード、検索条件等の入力フォームを作成して表示装置１に表示させたり、表示に対する入力装置２からのユーザの入力を受け付け、その入力から問い合わせ（クエリ）言語を作成してサーチ

7

エンジン４に渡す。

【００１２】サーチエンジン４は渡された問い合わせ言語により、構造化文書が蓄積されているドキュメントDB（データベース）７に対して検索を行う。SGMLパーサ５はDTD を基に、SGML文書を構文解析（パース）してその妥当性を検証し、SGML文書のエンティティツリーを作成する。

【００１３】エンティティマネージャ６は構文解析した結果をドキュメントDB７へ格納し、文書の実体を管理する。ドキュメントDB７には、ドキュメントを構成するエレメントの名称及びエレメントの階層構造がDTD により定義された構造化文書がエレメントに分割されて蓄積されている。上述のSGMLパーサ５、エンティティマネージャ６、及びドキュメントDB７によって構成されるSGMLシステムは、構造化文書を構文解析し、エレメントに分割して格納する機能を提供する。

【００１４】スタイルシートDB９は、SGML文書のDTD に応じた文書成形の規則を定めたスタイルシートが蓄積されている。DSSSL エンジン８は、構文解析されたSGML文書を、スタイルシートDB９に蓄積されている、SGML文書のDTD に応じた文書成形の規則を定めたスタイルシートに沿って表示するための文書成形を行う。上述のDSSSL エンジン８及びスタイルシートDB９によって構成されるDSSSL システムは、SGML文書を表示するための機能を提供する。

【００１５】インタフェースモジュール３は、SGMLシステム(5,6,7) から文書構造を読み出し、表示形式に沿って表示装置１に入力フォームを出力し、またDSSSL システム(8,9) と連携してキーワードを取得する。サーチエンジン４はSGMLシステム(5,6,7) 及びDSSSL システム(8,9) と連携して、与えられたクエリに応じて検索を行う。

【００１６】本発明の構造化文書データベース検索方法、構造化文書検索システム及び記録媒体は、インタフェースモジュール３が、ドキュメントDB７に蓄積されているDTD 、及びスタイルシートDB９に蓄積されているそのDTD のスタイルシートによって、エレメントの内容のレイアウト位置を規定した文書フォーマット図を、表示装置１に、例えば枠組みで表示し、文書フォーマット図のレイアウト位置の上下左右等、レイアウト位置に対応付けたキーワードの入力を受け付け、キーワードが入力されたレイアウト位置によって検索対象のエレメントを識別する。サーチエンジン４は、入力されたキーワードに基づいて、ドキュメントDB７に蓄積されているドキュメントの中の、インタフェースモジュール３が識別したエレメントの内容を検索し、文書を獲得する。

【００１７】また、本発明の構造化文書データベース検索方法、構造化文書検索システム及び記録媒体は、インタフェースモジュール３が、ドキュメントDB７に蓄積されているDTD によって、まずエレメントの階層構造のうち、所定の階層に属するエレメント名とその階層関係を表す情報とを表示装置１に表示し、表示したエレメント名及びその階層関係を表す情報とに基づく、エレメント名を表示すべき階層の指定を受け付け、指定された階層のエレメント名を、その階層関係を表す情報とともに表示装置１に表示し、エレメント名に対応付けた、例えばダイアログボックスへのキーワードの入力を受け付け、キーワードが入力されたエレメント名によって検索対象のエレメントを識別する。サーチエンジン４は、入力されたキーワードに基づいて、ドキュメントDTD ７に蓄積されているドキュメントの中の、インタフェースモジュール３が識別したエレメントの内容を検索し、文書を獲得する。

8

【００１８】さらに、本発明の構造化文書データベース検索方法、構造化文書検索システム及び記録媒体は、インタフェースモジュール３が、ドキュメントDB７に蓄積されているDTD によって、検索対象のエレメントを含む文書を構成する全てのエレメントのエレメント名を用いて、エレメントの階層構造を示す構造図を表示し、この構造図のエレメント名に対応付けた、例えばダイアログボックスへのキーワードの入力を受け付け、キーワードが入力されたエレメント名によって検索対象のエレメントを識別する。サーチエンジン４は、入力されたキーワードに基づいて、ドキュメントDTD ７に蓄積されているドキュメントの中の、インタフェースモジュール３が識別したエレメントの内容を検索し、文書を獲得する。

【００１９】従って、本発明の構造化文書データベース検索方法、構造化文書検索システム及び記録媒体は、文書構造がユーザにとって未知である構造化文書を検索する場合であっても、検索対象のエレメントの指定、及びそのエレメントに対するキーワードの入力が容易であって、検索を短時間で行うことができる。

【００２０】

【発明の実施の形態】以下に、本発明の構造化文書データベース検索方法（以下、本発明方法という）を図に基づいて説明する。まず、本発明方法に、前述の文書フォーマット図であるスタイル情報を用いる場合について、図２のキーワード入力の説明図、図３のフローチャート、及び図４のキーワード入力例の図に基づき説明する。

【００２１】SGMLでは文書型定義(DTD) で文書の型、即ち、章、節、項といった文書構造を定義する。インタフェースモジュール３は、ドキュメントDB７に格納されているSGML文書のDTD を読み出し、「レポート」「手紙」といった、ツリー構造の最上位の階層のエレメントを一覧表示する。

【００２２】ユーザは、画面に一覧表示されたDTD の中から、検索する文書の型に合ったDTD として、例えば「レポート」を選択して、スタイル情報を表示すべきDTD を決定する（ステップＳ１）。DSSSL エンジン８は、

9

決定されたDTD 、本例では「レポート」の文書成形の規則を定めたスタイルシートをスタイルシートDB９から読み込み（ステップＳ２）、フローオブジェクトを生成して、図２に示すような「レポート」の文書フォーマット図を表示装置１に表示する（ステップＳ３）。

【００２３】この文書フォーマット図により、ユーザはその文書構造を知らなくても、「レポート」が、「宛先」「発信元」「タイトル」「本文」及び「担当」のエレメントから構成され、「本文」が「概要」及び「詳細」のエレメントから構成されていることと、それぞれのエレメントの内容の表示されるレイアウト位置がどの辺りであるかを、その枠組みで視覚的に把握することができる。

【００２４】インタフェースモジュール３は、表示された文書フォーマット図に対してキーワードの入力があるか否かを監視し（ステップＳ４）、キーワードが入力された場合は、その入力位置から、キーワードが入力されたエレメントを判定する。インタフェースモジュール３は、入力されたキーワードからクエリを作成してサーチエンジン４に渡す（ステップＳ５）。サーチエンジン４はインタフェースモジュール３から渡されたクエリにより、ドキュメントDB７の中の検索対象のエレメントの内容を検索して目的の文書を獲得する（ステップＳ６）。

【００２５】このとき、図４に示すように、複数のエレメントに対するキーワードの入力が可能である。複数のエレメントに対してキーワードが入力された場合、インタフェースモジュール３は、入力された複数のキーワードとデフォルトで設定されている検索条件（OR検索又はAND 検索）とからクエリを作成するか、又は、図４に示すように、検索条件の入力画面を表示してユーザに選択を促し、選択された検索条件と入力されたキーワードとからクエリを作成する。

【００２６】例えば、「タイトル」エレメントにキーワード「見積もり」が、また「担当」エレメントにキーワード「富士太郎」が入力され、AND 検索を行う場合は、「タイトル」エレメントの中に文字列「見積もり」が含まれるとともに、「担当」エレメントの中に文字列「富士太郎」が含まれる文書を検索する。上述の入力でOR検索を行う場合は、エレメント「タイトル」の中に文字列「見積もり」が含まれるか、又はエレメント「担当」の中に文字列「富士太郎」が含まれる文書を検索する。

【００２７】図４は、検索条件の指定においてエレメント間の関係(and , or) を表したリンクを別のリンクの対象とする場合の入力例を示している。「宛先」エレメントにキーワード「経理」、「発信元」エレメントにキーワード「経理」、「タイトル」エレメントにキーワード「見積もり」、「担当」エレメントにキーワード「富士太郎」が入力されている。

【００２８】入力したキーワードによって、「AND 検索」「OR検索」「複合検索」のいずれを実行するかをユーザに選択させるために、例えばボタン、メニュー等を表示する。さらに「複合検索」においては「AND リンク」「ORリンク」のいずれを設定するかをユーザに選択させるためのボタン、メニュー等を表示する。

10

【００２９】例えば、ユーザが、ボタン、メニュー等の「複合検索」及び「AND リンク」を選択し、ポインタＰによって「発信元」、「タイトル」、及び「担当」エレメントを指定すると、これらのエレメントが、例えば実線で結ばれ、AND リンクが設定されたことが表示される。次に、ユーザが「ORリンク」を選択し、AND リンクの実線と、「宛先」エレメントをポインタＰによって指定すると、AND リンクの実線と「宛先」エレメントとが、例えば破線で結ばれ、ORリンクが設定されたことが表示される。

【００３０】以上の設定の結果、「発信元」エレメントに文字列「経理」が含まれ、かつ「タイトル」エレメントに文字列「見積もり」が含まれ、かつ「担当」エレメントに文字列「富士太郎」が含まれる文書か、又は「宛先」エレメントに文字列「経理」が含まれる文書を検索するクエリが作成される。

【００３１】さらに、検索条件として、キーワードが入力されたエレメントだけを検索対象とする「単体検索」か、そのエレメントの下位の階層のエレメントも検索対象に含む「階層検索」かを選択することもできるようになっている。

【００３２】なお、検索条件の選択方法はメニュー、ボタンに限定されるものではなく、文字入力、番号選択等の他の選択方法も可能である。また、複合検索におけるAND リンク、ORリンクの別を表す方法は実線、破線等の線種による方法に限らず、線の色を変える、AND/ORの文字で表示する等の他の方法であってもよい。さらに、必ずしも表示しなくてもよい。

【００３３】次に、本発明方法にエレメント構造情報のみを用いる場合について、図５のキーワード入力の説明図、図６のフローチャート、及び図７のキーワード入力例の図に基づいて説明する。前述の場合と同様に、インタフェースモジュール３は、ドキュメントDB７に格納されているSGML文書のDTD を読み出し、「レポート」「手紙」といった、ツリー構造の最上位の階層の一覧を表示するが、本例ではリスト形式で表示する（ステップＳ１１）。

【００３４】ユーザは、画面にリスト形式で一覧表示されたDTD の中から、検索する文書の型に合ったDTD を決定する（ステップＳ12）。このとき、インタフェースモジュール３は、例えば「レポート」の文字列上がクリックされたか、又は「レポート」の欄に設けられた、下位の階層が存在することを示すマークがクリックされたかを識別して、エレメントが選択されたか否かを判定する（ステップＳ13）。エレメントの欄のマークがクリックされた場合、「レポート」の下位の階層の「宛先」「タ

11

イトル」…のエレメント名（タグ名）をリスト形式で表示する（ステップＳ14）。

【００３５】「レポート」の階層の場合と同様、下位の階層のエレメントの文字列上がクリックされたか、又はエレメントの欄に設けられているマークがクリックされたかを識別して、エレメントが選択されたか否かを判定する（ステップＳ13）。例えば、「本文」のエレメントの欄のマークがクリックされた場合は、さらに下位の階層のタグリストを表示する（ステップＳ14）。

【００３６】一方、エレメントの文字列がクリックされた場合は、キーワードの入力画面を表示し、カーソルを点滅させてキーワードの入力待ちとなる（ステップＳ15）。キーワードが入力されると、検索の開始が指示されたか否かをチェックし（ステップＳ16）、検索の開始が指示されない場合はステップＳ13に戻って、ユーザからの次の入力待ちになる。

【００３７】前述の場合と同様、入力したキーワードによって、「AND 検索」「OR検索」「複合検索」のいずれを実行するかをユーザに選択させるために、例えばボタン、メニュー等を表示する。さらに「複合検索」においては「AND リンク」「ORリンク」のいずれを設定するかをユーザに選択させるためのボタン、メニュー等を表示する。また、前述の場合と同様、検索条件として、キーワードが入力されたエレメントだけを検索対象とする「単体検索」か、そのエレメントの下位の階層のエレメントも検索対象に含む「階層検索」かを選択することもできるようになっている。

【００３８】検索の開始が指示されると、インタフェースモジュール３は、入力されたキーワードからクエリを作成して（ステップＳ17）サーチエンジン４に渡し、サーチエンジン４はインタフェースモジュール３から渡されたクエリにより、ドキュメントDB７の中の検索対象のエレメントの内容を検索して目的の文書を獲得する（ステップＳ18）。

【００３９】このとき、図７に示すように、複数のエレメントに対するキーワードの入力が可能である。複数のエレメントに対してキーワードが入力された場合、インタフェースモジュール３は、入力された複数のキーワードとデフォルトで設定されている検索条件（OR検索又はAND 検索）とからクエリを作成するか、又は、図７に示すように、検索条件の入力画面を表示してユーザに選択を促し、選択された検索条件と入力されたキーワードとからクエリを作成する。

【００４０】図７のようなキーワード入力で、AND 検索と単体検索とが選択された場合、「レポート」エレメントに文字列「経理」が含まれ、かつ「タイトル」エレメントに文字列「見積もり」が含まれ、かつ「本文」エレメントに文字列「FMV 」が含まれる文書を検索するクエリが作成される。

【００４１】また図７のようなキーワード入力で、OR検索と単体検索とが選択された場合、「レポート」エレメントに文字列「経理」が含まれるか、又は「タイトル」エレメントに文字列「見積もり」が含まれるか、又は「本文」エレメントに文字列「FMV 」が含まれる文書を検索するクエリが作成される。なお、各エレメントに与えるキーワードにAND 、OR、正規表現を記述することもできる。

12

【００４２】図７のようなキーワード入力で、AND 検索と階層検索とが選択された場合、「レポート」エレメントより下位のエレメントに文字列「経理」が含まれ、かつ「タイトル」エレメントに文字列「見積もり」が含まれ、かつ「本文」エレメントより下位のエレメントに文字列「FMV 」が含まれる文書を検索するクエリが作成される。

【００４３】以上のように、本例の場合は階層順を追ってエレメントの構造図を表示していくので、一画面でエレメントの構造を表示できない巨大な構造の構造化文書に対しても、目的のエレメントにキーワードをスムーズに与えることができる。

【００４４】次に、本発明方法にエレメント構造情報のみを用いる他の場合について、図８のキーワード入力の説明図、図９のフローチャート、及び図10のキーワード入力例の図に基づいて説明する。前述の場合と同様に、インタフェースモジュール３は、ドキュメントDB７に格納されているSGML文書のDTD を読み出し「レポート」「手紙」といった、ツリー構造の最上位の階層のエレメントを一覧表示する。

【００４５】ユーザは、画面にリスト形式で一覧表示されたDTD の中から、検索する文書の型に合ったDTD を決定する（ステップＳ21）。DSSSL エンジン８は、決定されたDTD 、本例では「レポート」の文書成形の規則を定めたスタイルシートをスタイルシートDB９から読み込み、フローオブジェクトを生成して、図８に示すような、「レポート」のツリー構造の構造図を表示装置１に表示する（ステップＳ22）。

【００４６】この構造図により、ユーザはその文書構造を知らなくても、「レポート」が、「宛先」「発信元」「タイトル」「本文」及び「担当」のエレメントから構成され、「本文」が「概要」及び「詳細」のエレメントから構成されていることを把握することができる。

【００４７】インタフェースモジュール３は検索の開始が指示されたか否かをチェックし（ステップＳ23）、検索開始が指示されない場合は、表示した構造図のエレメントに対する選択入力待ちとなる（ステップＳ24）。いずれかのエレメントが選択されると、そのエレメントに対応付けて、キーワードの入力画面を表示し、カーソルを点滅させてキーワードの入力待ちとなる（ステップＳ25）。

【００４８】キーワードが入力されると、検索の開始が指示されたか否かをチェックし（ステップＳ23）、検索

13

の開始が指示されない場合は、表示した構造図のエレメントに対する選択入力待ちとなる（ステップＳ24）。いずれかのエレメントが選択されると、そのエレメントに対応付けて、キーワードの入力画面を表示し、カーソルを点滅させてキーワードの入力待ちとなる（ステップＳ25）。

【００４９】検索の開始が指示されると、インタフェースモジュール３は、入力されたキーワードからクエリを作成してサーチエンジン４に渡し（ステップＳ26）、サーチエンジン４はインタフェースモジュール３から渡されたクエリにより、ドキュメントDB７の中の検索対象のエレメントの内容を検索して目的の文書を獲得する（ステップＳ27）。

【００５０】前述の場合と同様、入力したキーワードによって、「AND 検索」「OR検索」「複合検索」のいずれを実行するかをユーザに選択させるために、例えばボタン、メニュー等を表示する。さらに「複合検索」においては「AND リンク」「ORリンク」のいずれを設定するかをユーザに選択させるためのボタン、メニュー等を表示する。

【００５１】また、前述の場合と同様、検索条件として、キーワードが入力されたエレメントだけを検索対象とする「単体検索」か、そのエレメントの下位の階層のエレメントも検索対象に含む「階層検索」かを選択することもできるようになっている。なお、各エレメントに与えるキーワードにAND 、OR、正規表現を記述することができる。

【００５２】このとき、図10に示すように、複数のエレメントに対するキーワードの入力が可能である。複数のエレメントに対してキーワードが入力された場合、インタフェースモジュール３は、入力された複数のキーワードとデフォルトで設定されている検索条件（OR検索又はAND 検索）とからクエリを作成するか、又は、図10に示すように、検索条件の入力画面を表示してユーザに選択を促し、選択された検索条件と入力されたキーワードとからクエリを作成する。

【００５３】図10のようなキーワード入力で、AND 検索と単体検索とが選択された場合、「タイトル」エレメントに文字列「見積もり」が含まれ、かつ「担当」エレメントに文字列「富士太郎」が含まれる文書を検索するクエリが作成される。また図10のようなキーワード入力で、OR検索と単体検索とが選択された場合、「タイトル」エレメントに文字列「見積もり」が含まれるか、又は「担当」エレメントに文字列「富士太郎」が含まれる

14

文書を検索するクエリが作成される。

【００５４】なお、上述のようなキーワード入力のいずれかをユーザに選択させるメニュー画面を設けてもよい。

【００５５】

【発明の効果】以上のように、本発明の構造化文書データベースの検索方法、検索システム及び記録媒体は、構造化文書の文書構造の図式をエレメント名を用いてユーザに示し、図式上のエレメント名に対してユーザがキーワードを入力したエレメントを対象として検索するので、文書構造がユーザにとって未知である構造化文書を検索する場合であっても、検索対象のエレメントの指定、及びそのエレメントに対するキーワードの入力が容易であって、検索が短時間で行えるという優れた効果を奏する。

【図面の簡単な説明】

【図１】本発明のシステム構成図である。

【図２】スタイル情報を用いたキーワード入力の説明図である。

【図３】図２による検索方法のフローチャートである。

【図４】スタイル情報を用いたキーワード入力例の図である。

【図５】構造情報を用いたキーワード入力の説明図（その１）である。

【図６】図５による検索方法のフローチャートである。

【図７】図５の構造情報を用いたキーワード入力例の図である。

【図８】構造情報を用いたキーワード入力の説明図（その２）である。

【図９】図８による検索方法のフローチャートである。

【図１０】図８の構造情報を用いたキーワード入力例の図である。

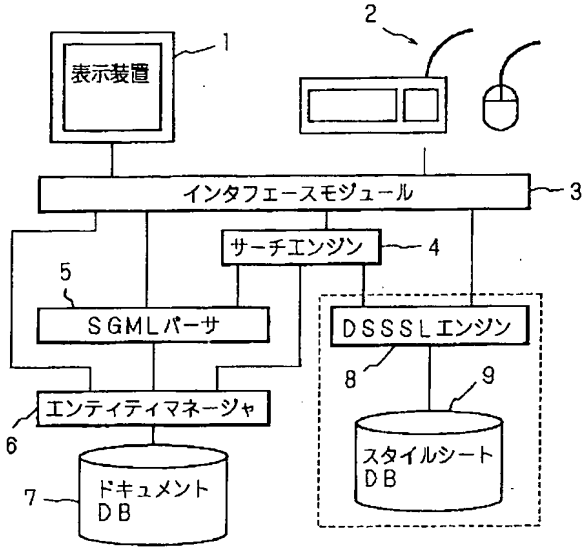
【図１１】構造化文書の構造の一例の図である。

【符号の説明】

1　表示装置
2　入力装置
3　インタフェースモジュール
4　サーチエンジン
5　SGMLパーサ
6　エンティティマネージャ
7　ドキュメントDB
8　DSSSL エンジン
9　スタイルシートDB

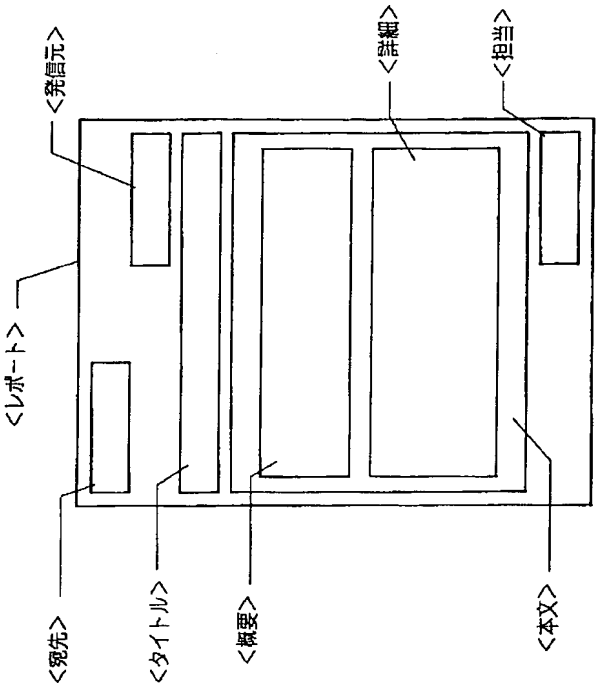
【図１】

本発明のシステム構成図



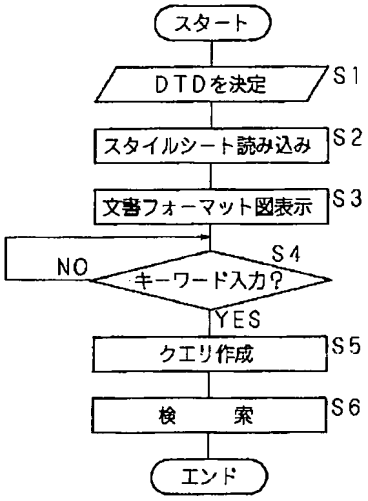
【図２】

スタイル情報を用いたキーワード入力の説明図



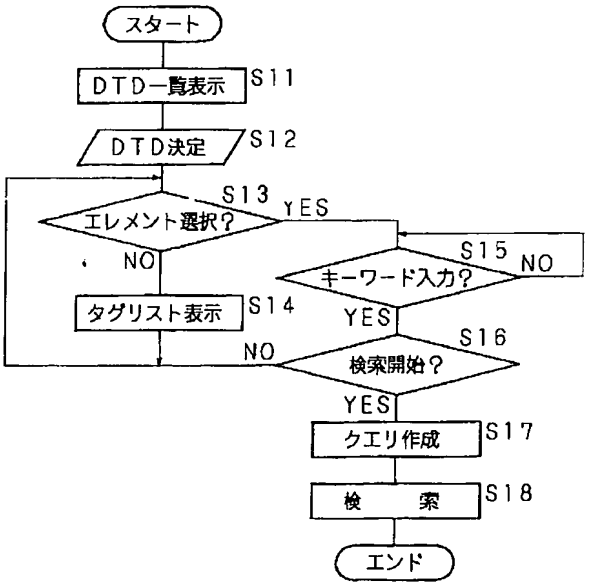
【図３】

図2による検索方法のフローチャート



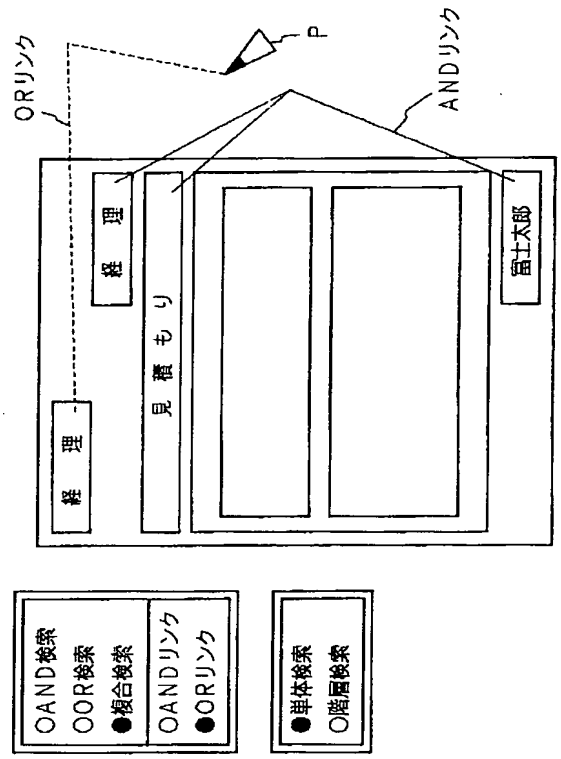
【図６】

図5による検索方法のフローチャート



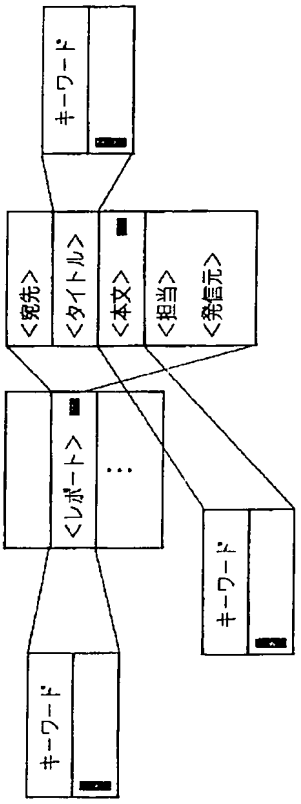
【図４】

スタイル情報を用いたキーワード入力例の図



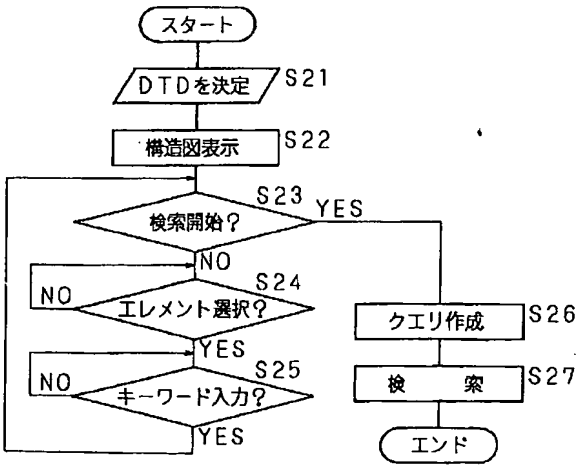
【図５】

構造情報を用いたキーワード入力の説明図（その１）



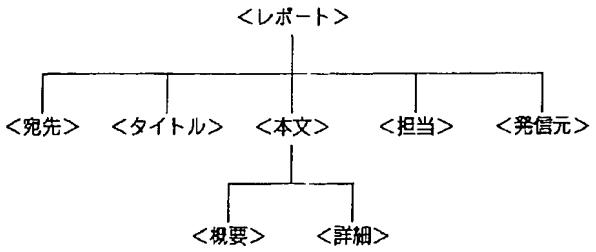
【図９】

図８による検索方法のフローチャート



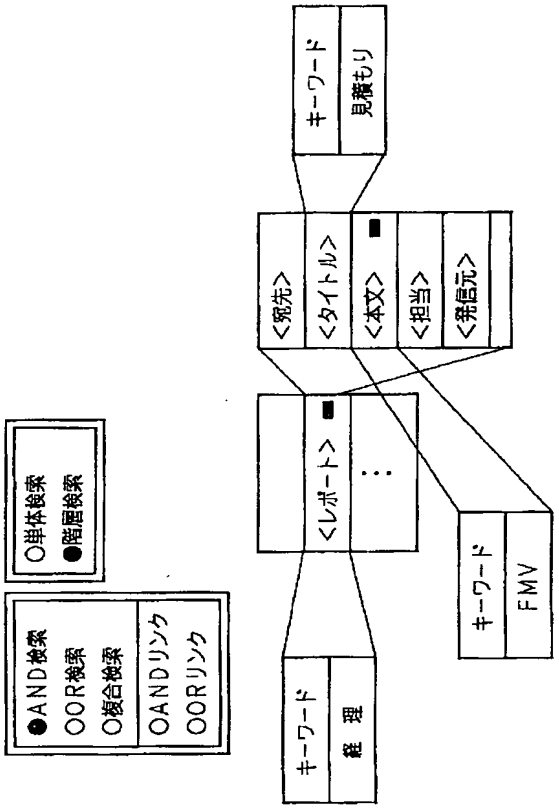
【図１１】

構造化文書の構造の一例の図



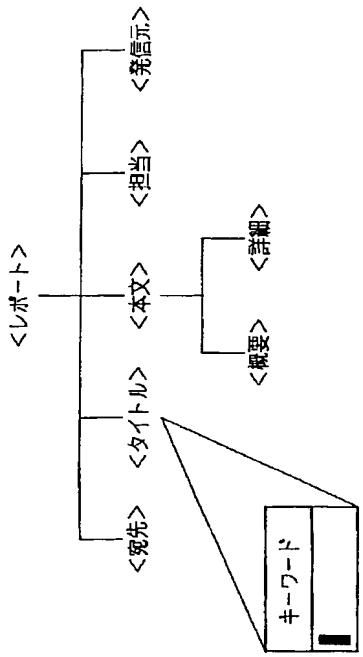
【図７】

図５の構造情報を用いたキーワード入力例の図



【図８】

構造情報を用いたキーワード入力の説明図（その２）



【図１０】

図８の構造情報を用いたキーワード入力例の図

